**Testimony of**
**Congressman Tim Murphy**
**Pennsylvania – 18th**
**Before the U.S. House Committee on Education and Labor**
**Reauthorization of No Child Left Behind**
**Wednesday, May 16, 2007**

Thank you Mr. Chairman for asking for input regarding reauthorization and recommendations for NCLB. I'm going to limit my comments today to one area where I believe some changes can have a dramatic and very helpful impact.

As I listen to comments from parents, teachers and school administrators about No Child Left Behind, one common theme they discuss involves the way tests are used and interpreted. In over 25 years of practice as a psychologist, I have administered thousands of tests to school children including IQ tests, academic achievement tests, aptitude, to psychological measures of personality or behavior.

My psychology training always emphasized the tests must meet very specific criteria of being considered scientifically reliable and valid. By "valid" I mean a test should measure what it is supposed to measure (such as reading level or intelligence), and by "reliable" I mean a test should provide consistent information when administered to different children at different schools. Of the many tests for NCLB administered across this nation, the tests must meet these rigorous standards. If they do, then they will have some meaning to parents, teachers and students. If they do not, then the scores are fairly worthless.

A few years ago a doctor analyzed the school achievement results for an exam used statewide. He found the test was unable to separate out how kids were really doing. Oddly enough, he found that every child was scoring pretty well. He labeled this the "Lake Woebegone Effect" after the popular radio program, since all the kids indeed were above average. When states lower standards so everyone passes, the tests are scientifically worthless.

Next, when it comes to interpreting tests scores, it is absolutely essential that the scores are interpreted only within a specific purpose for that test. For example, scores on a math achievement test may tell you how a student is doing in mastery of math skills at that time. It should not be used to interpret how a child will do in college. And you shouldn't interpret a child's growth based on a test he or she did not take.

With the tests use for NCLB, that I believe we are actually interpreting the test results in a way that in most states may not be scientifically valid. Under NCLB, states set annual proficiency targets to determine how students are doing at that point in time. This is generally referred to as a "status model". However, when these scores are used to determine if a school is making progress (known as the Adequate Yearly Progress or AYP), I believe it is a quantum leap that stretches interpretation in ways that are not fully supported by the data.

For example, if you want to know how a high school football team is doing, you can look at its overall win/loss record. A team in 2005 that has 9 wins and 0 losses may look pretty good. Now let's assume that same team in 2006 dropped to 4 wins and 5 losses. By using the AYP model we would conclude the 2006 team and coach are failures. However, if you take into account that the undefeated team may have been made up of all senior students who then graduated, the next year's team was less experienced, and the other schools teams were also made up of entirely different players, then you could not so easily conclude the coach was a failure.

With the test interpretation used in NCLB, when you compare a fifth grade in 2004 to a fifth grade in 2005, to a fifth grade in 2006, you are looking at groups of entirely different students. One year the class may have several gifted students in it. The next year there may be several students with learning problems. In each case the composition of the class will skew the scores. The gifted students may add several points to the score one year, and the next year the scores may drop down, all independent of the teacher's abilities.

Further, when you interpret the results based upon where a student falls in comparison with the "average" of the previous year's students, you really miss out on obtaining valuable data for individual students. We can say if a classroom average score may have gone up or down, but we really don't know if the individual students improved or not.

So what do we do?

Some states have started to use a Growth Model, to measure student performance over time. In 2005 Secretary Spellings began a pilot program to test out the growth model. Eventually ten states will be enrolled in this program allowing schools to track student progress rather than just take static measures of their grade. I believe the Secretary is on the right track. After all, as a parent or a teacher, would you like to know how your child's class compares to classes in the past? Or would you like to know how your own child is progressing? Would you like to know if the class is doing better at mastery of reading and math skills? Or would you like to know how your own child has gained ground in reading and math? Using growth models where you can track each student improvements over time, you would be able to answer all of these questions.

My recommendation would be that schools throughout the nation shift their focus to gathering growth model data that is more meaningful for teachers, parents and students.

With a Growth model, we will be able to see if even the most academically challenged student is gaining or losing ground each year. This allows schools to focus on helping out each child at their own ability level and pace.

In the context of fair testing, let me add that testing must be done at the students level. It is unfair, and dare I say, it borders on being abusive when we demand an eighth grade child with $4^{th}$ grade reading skills, take tests with questions worded on the eighth grade level. Similarly, if sophomore high school student with reading problems is given a fourth grade test that meets their reading level but is totally inappropriate on social and emotional content, you run the risk of alienating the student from the test and thus throwing off the reliability and validity of the scores.

Finally, we need to understand, as difficult as it may be to admit, that some children simply cannot perform at 100% level for their age and grade. Children with severe learning problems, or substantial developmental disorders may never be able to match their classmates. If schooling was all it took to overcome these limits, that would be wonderful. However, if we judge all children by the same standard, we leave many students with the only possibility of failure. By using tests that better adapt to the child's abilities, and then track them over time, we will indeed know whether or not that child is making progress. Isn't that what we really want to know?

This is an ideal time to make changes in No Child Left Behind. Let's learn from errors of the past, add in sound scientific knowledge and make sure we are truly evaluating progress and problems. Every parent, teacher and administrator will benefit from having more accurate results. And we will have the information we need to make sure no child is left behind.